LINK PREDICTION ON SOCIAL MEDIA

# Himani Sheoran\*1, Udit Jain\*2, Umang Tiwari\*3 Ms. Savita Sharma\*4

\*1,\*2,\*3,Student, Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology , Delhi, India.

\*4Assistant Professor, Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, India.

# ABSTRACT

Currently, social networks have brought about an enormous number of users connecting to such systems over a couple of years, whereas the link mining is a key research track in this area. It has pulled the consideration of several analysts as a powerful system to be utilized as a part of social networks study to understand the relations between nodes in social circles. Numerous data sets of today’s interest are most appropriately called as a collection of interrelated linked objects. The main challenge faced by analysts is to tackle the problem of structured data sets among the objects. For this purpose, we design a new comprehensive model that involves link mining techniques with semantics to perform link mining on structured data sets. The past work, to our knowledge, has investigated on these structured datasets using this technique. For this purpose, we extracted real time data of posts using different tools from one of the famous SN platforms and check the society’s behavior against it. We have verified our model utilizing diverse classifiers and the derived outcomes inspiring.

**Keywords:** Machine Learning, Parkinson’s Disease, SMOTE, Hyperparameter Tuning, Xgboost, Catboost,

Random Forest Classifier.

# INTRODUCTION

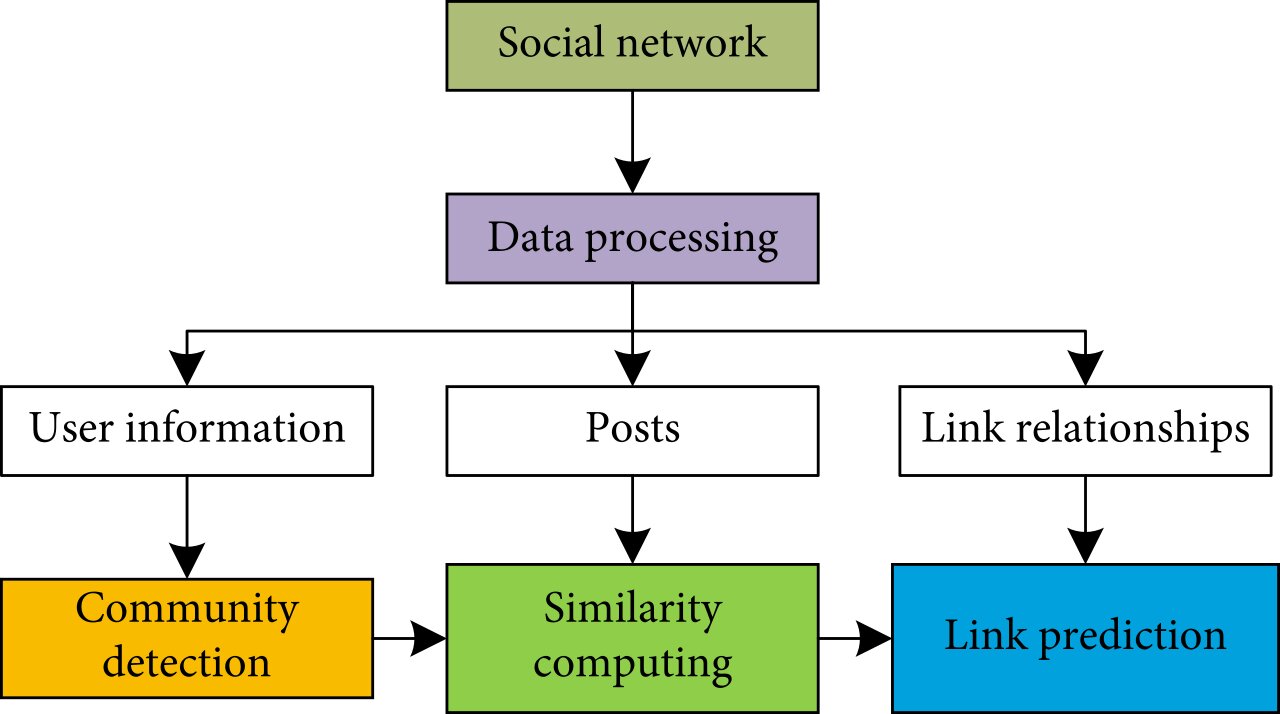
Link prediction is to predict whether there will be links between two nodes based on the attribute information and the observed existing link information. Link prediction not only can be used in the field of social network but can also be applied in other fields. As in bioinformatics, link prediction can be used to discover interactions between proteins; in the field of electronic commerce, link prediction can be used to create the recommendation system; and in the security field, link prediction can help to find the hidden terrorist criminal gangs. Link prediction is closely related to many areas. Therefore, in recent years there is a lot of correlation algorithms proposed to solve the problem of link prediction. Social networks are a popular way to interpret the interaction among the people. They can be

Visualized as graphs, where a vertex corresponds to a person and edges represent the connection between them. Understanding the dynamics that drive the evolution of social networks is a complex problem due to a large number of variable parameters. But, a comparatively easier problem is to understand the association between two specific nodes.

For the given source node and destination node we have to predict whether there is any probability of connecting between them.

# METHODOLOGY

Followings are the steps that has been taken to build the efficient model for early detection of Parkinson’s Disease:



## Dataset Detail:

-

## Data Preprocessing:

This step contains two process which is Normalization and balancing the dataset which is explain in detail below: **Normalization:**

Normalization is a technique which is applied as part of data preparation in machine learning. The need of normalization is to change the values of numeric columns in the dataset to a common scale, without changing differences in the ranges of values.

## Balance Dataset:

Balancing of the dataset is required when there is a class in minority which made the dataset made the dataset biased towards the other class. We balanced our data using SMOTE (Synthetic Minority Oversampling Technique) with the help of imbalance library.

**Models:**

# MODELING AND ANALYSIS

Following are the models that we used:

XGBOOST:

XGBoost is a gradient boosting library. It helps to implements machine learning algorithms under the Gradient Boosting framework. XGBoostis a parallel tree boosting which solves many Machine Learning problems in a fast and simple way. The same code runs on distributed environment and solves many machine learning problems.

## Hyperparameter Tuning:

The aim of hyperparameter tuning is to get the best possible parameter for our model. We did Hyperparameter tuning with the help of GridSearchCV cause it searches for best set of hyperparameters from a grid of hyperparameters values.

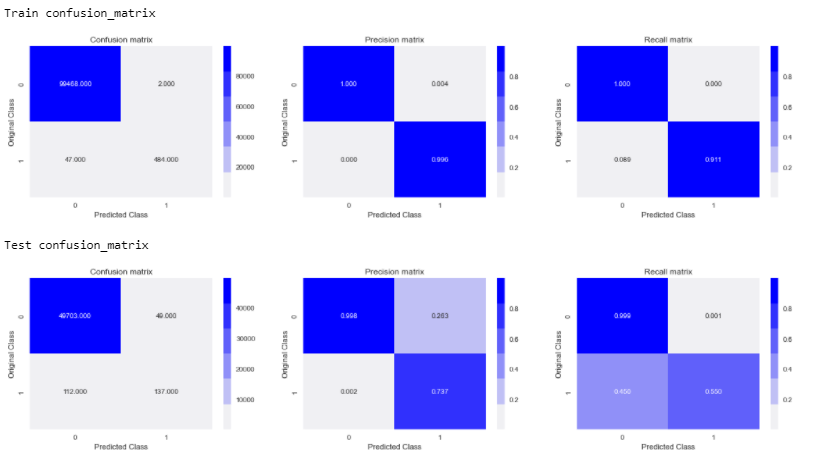
## MATHEW CORRELATION COFFECIENT (MCC):

The Matthews correlation coefficient (MCC) or phi coefficient is used to measure of the quality of binary classifications, introduced by biochemist Brian W. Matthews in 1975. The range of values of MCC lie between -1 to +1. MCC takes all the four value of values of confusion matrix into account. If the MCC value is close to 1 means that both classes are predicted well.

Where TP is the True Positive, TN is the True Negative, FP is the False Positive and FN is the False Negative.

# RESULTS AND DISCUSSION

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **F1-Score** | **Accuracy** |
| **XgBoost** | **0.62988505** | **99.67** |

****

\*Accuracy Comparison

\*MCC Comparison

# CONCLUSION

Early detection of Parkinson’s diseases is very useful as it will helps to prevent the patients from worst stage. From this study we analyses the different machine learning algorithm like XGBoost and Random Forest and got an efficient Parkinson’s Disease prediction model with high accuracy.

# REFERENCES

* Liyan Dong,1,2 Yongli Li,3 Han Yin,1,2 Huang Le,1,2 and Mao Rui1, The Algorithm of Link Prediction on Social Network, Volume 2013 |Article ID 125123
* [Mahdi Jalili](https://royalsocietypublishing.org/doi/10.1098/rsos.160863), [Yasin Orouskhani](https://royalsocietypublishing.org/doi/10.1098/rsos.160863), [Milad Asgari](https://royalsocietypublishing.org/doi/10.1098/rsos.160863), [Nazanin Alipourfard](https://royalsocietypublishing.org/doi/10.1098/rsos.160863) and [Matjaž Perc](https://royalsocietypublishing.org/doi/10.1098/rsos.160863), Link prediction in multiplex online social networks, Published:01 February 2017 <https://doi.org/10.1098/rsos.160863>
* [Journal of the American Society for Information Science and Technology](https://asistdl.onlinelibrary.wiley.com/journal/15322890),The link-prediction problem for social networks,[David Liben-Nowell](https://asistdl.onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Liben-Nowell%2C+David),[Jon Kleinberg](https://asistdl.onlinelibrary.wiley.com/action/doSearch?ContribAuthorRaw=Kleinberg%2C+Jon), First published: 26 March 2007, <https://doi.org/10.1002/asi.20591>

# New perspectives and methods in link prediction,Ryan N. Lichtenwalter, Jake T. Lussier, [KDD '10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining](https://dl.acm.org/doi/proceedings/10.1145/1835804)July 2010 Pages 243-252 <https://doi.org/10.1145/1835804.1835837>

# Link Prediction using Supervised Learning ∗ Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki Rensselaer Polytechnic Institute, Troy, New York 12180 {alhasan, chaojv, salems, zaki}@cs.rpi.edu